

Intro til statistik

IV Analyse af antalstabeller

Claus Thorn Ekstrøm

Biostatistik, KU

ekstrom@sund.ku.dk

Onsdag 20. maj 2020

Slides @ biostatistics.dk/puff/



Plan for i dag

- Kategoriske outcomes
- Effektmål for risici

Titanic

```
data(Titanic) # Hent data
library("tidyverse")
as.data.frame(Titanic) %>% # Lav til data.frame
  group_by(Sex, Survived) %>% # Gruppér for hver kombination
  summarise(n = sum(Freq)) # Optæl antal observationer
```

```
## # A tibble: 4 x 3
## # Groups:   Sex [2]
##   Sex      Survived     n
##   <fct>   <fct>     <dbl>
## 1 Male     No         1364
## 2 Male     Yes         367
## 3 Female  No          126
## 4 Female  Yes         344
```

Binære udfald

For binære udfald kan vi ikke direkte regne.

S F F S S S S F S F F S S F S S S S F
1 0 0 1 1 1 1 0 1 0 0 1 1 0 1 1 1 1 0

$$\hat{p} = \frac{12}{19} = 0.63$$

Estimatet er "bare" et gennemsnit.

Konfidensintervaller for andele

Estimatet for p (kaldet \hat{p}) udregnes som den observerede relative frekvens.

Hvordan laver vi konfidensintervaller for andele?

Et 95%-konfidensinterval for en **andel** p er givet ved ca.

$$[\hat{p} - 2 \cdot SE(\hat{p}); \hat{p} + 2 \cdot SE(\hat{p})]$$

I R bruger vi `prop.test()` med to argumenter: antal succeser og totalen.

Kategorisk forklarende variabel

```
## # A tibble: 4 x 3
## # Groups:   Sex [2]
##   Sex      Survived     n
##   <fct>   <fct>         <dbl>
## 1 Male    No             1364
## 2 Male    Yes             367
## 3 Female No             126
## 4 Female Yes             344
```

Er der forskel på overlevelsessandsynligheden?

$$H_0 : p_f = p_m$$

Effektmål for andele

Kan beskrive effekten på forskjellig skala:

- Risikodifferens
- Relativ risiko (risikoratio)
- Odds ratio

Sammenlign to grupper:

Hvor stor er en forskel på 1% og 5%? 51% og 55%?

Risikodifferens

Også kaldet absolut risikodifferens.

Man modellerer typisk *forskellen* mellem sandsynligheder.

$$RD = p_1 - p_2$$

Det svarer altså til en ændring i **procentpoint**.

Procentpoint kan dække over forholdsmæssige store og små ændringer.
Hvilken værdi svarer til ingen effekt?

Relativ risiko

En anden mulighed er at beskrive *forholdet* mellem sandsynligheder.

$$RR = \frac{p_1}{p_2}$$

Det svarer altså til et *forhold* mellem to **sandsynligheder**.

Relativ risiko kan dække over forskelsmæssige store og små ændringer.
Hvilken værdi svarer til ingen effekt?

Odds

Odds for en hændelse A er

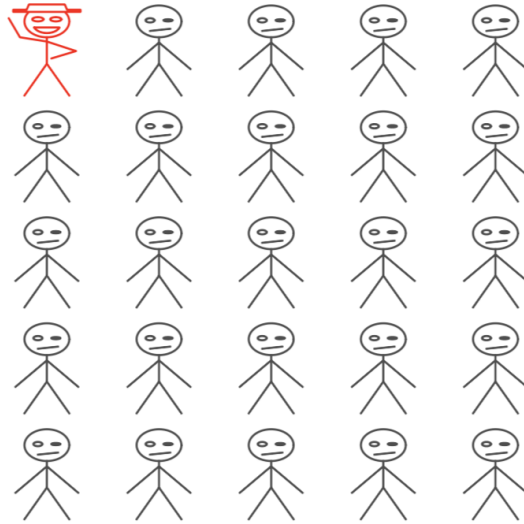
$$\text{Odds}(A) = \frac{\# \text{personer der oplever } A}{\# \text{personer der ikke oplever } A}$$

Sammenlign med risikobegrebet

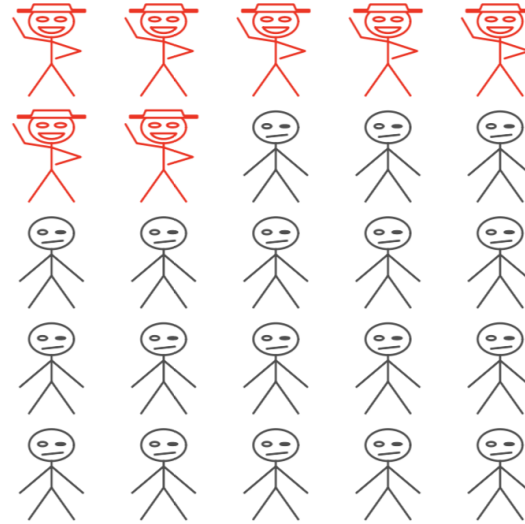
$$\text{Risiko}(A) = \frac{\# \text{personer der oplever } A}{\text{total antal personer}}$$

Vigtigt at skelne mellem odds og risiko - det er ikke det samme!

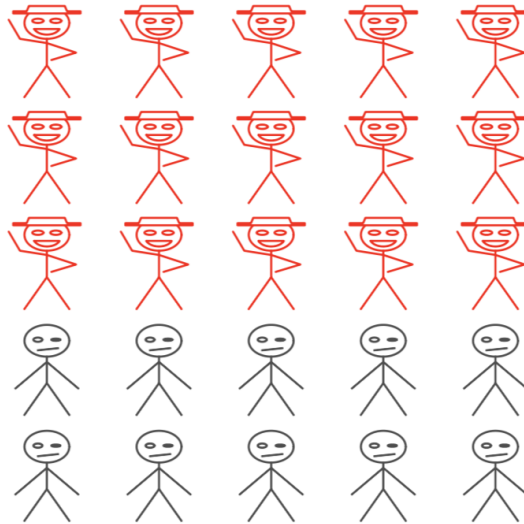
risiko = 1/25 (4.0%), odds = 1/24 (0.04)



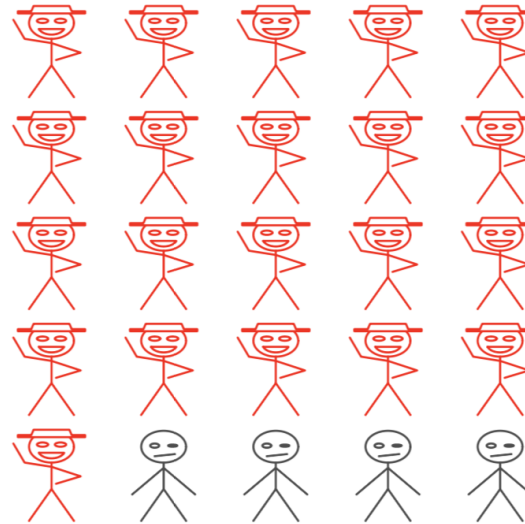
risiko = 7/25 (28.0%), odds = 7/18 (0.38)



risiko = 15/25 (60.0%), odds = 15/10 (1.5)



risiko = 21/25 (84.0%), odds = 21/4 (5.25)



Odds-ratio

Odds-ratio er forholdet mellem to odds, fx. odds fra to grupper:

$$\text{Odds}(A \mid \text{grp 1}) = \frac{\text{\#personer der oplever } A \text{ i grp 1}}{\text{\#personer der ikke oplever } A \text{ grp 1}}$$

$$\text{Odds}(A \mid \text{grp 2}) = \frac{\text{\#personer der oplever } A \text{ i grp 2}}{\text{\#personer der ikke oplever } A \text{ grp 2}}$$

$$\text{OR} = \frac{\text{Odds}(A \mid \text{gruppe 1})}{\text{Odds}(A \mid \text{gruppe 2})}$$

Hvad skal man bruge?

Hvordan tror vi effekter er? Hvilken skala?

- Relativ risiko kan kun korrekt udregnes i nogle designs (fx. prospektive kohortestudier), og har en intuitiv fortolkning.
- Odds ratio kan altid udregnes - også for case-control-forsøg --- og har nogle pæne matematiske egenskaber.
- Når forekomsten er lille (p lille) er RR og OR tæt på hinanden. Ellers ikke.
- Ofte bør man rapportere *både* den absolutte og den relative effekt.

Hvordan gør man det i R?

- `prop.test()` - en stikprøve for andel
- `matrix()` - lav matrix/tabel
- `chisq.test()` - test for ens fordeling
- `fisher.test()` - KI for OR
- `monte_carlo_chisq_test()` fra MESS - simuleret χ^2 -test med faste marginaler.
- `riskratio()` fra `epitools` - KI for RR